AD_____

Award Number: W81XWH-09-1-0721

TITLE: Novel Strategy for Lineage Tracing of Cancer Stem Cells by Lentiviral Barcoding

PRINCIPAL INVESTIGATOR: Edmund Chang

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, TX 77030

REPORT DATE: October 2012

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
October 2012	Annual Summary	15 September 2009 - 14 September 2012

4. TITLE AND SUBTITLE
Novel Strategy for Lineage Tracing of Cancer Stem Cells by Lentiviral Barcoding

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0721

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Edmund Chang

E-Mail: ec1@bcm.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Baylor College of Medicine
Houston, TX 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The cancer stem cell theory posits a subpopulation of tumorigenic cells with self-renewal potential and the capability to wholly reconstitute original tumors by lineage restriction. We proposed a novel lentiviral tagging system to genetically trace immunosorted cell fate, and further investigate relative TIC enrichment through immunosorting protocols. We have developed a genetic-based strategy to dramatically improve sensitivity and experimental time in prospective identification of novel biomarkers. In months 25-36, we build on our previous results and implemented next-generation sequencing methodology to further trave the tumor sublineages in the mouse mammary tumor model.

15. SUBJECT TERMS
Cancer stem cell, lineage tracing, lentiviral barcoding

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	10	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

Introduction

The cancer stem cell theory posits a subpopulation of tumorigenic cells with self-renewal potential and the capability to wholly reconstitute original tumors by lineage restriction. Though surface markers such as CD29/CD24 in mice and CD44/CD24 in humans have facilitated enrichment of tumor-initiating cells (TICs), no lineage-tracing experiments have been performed to date on TIC cell fate. We proposed a novel lentiviral tagging system to genetically trace immunosorted cell fate, and further investigate relative TIC enrichment through immunosorting protocols. Using the well-characterized p53-null mouse mammary tumor model, we demonstrated that it was feasible to determine TIC population for the CD29/CD24 biomarker combination. Genetic-based quantitation dramatically improved the sensitivity, statistical robustness and experimental time in prospective identification of novel biomarkers. I will list and discuss the milestones reached in months 25-36 of the fellowship in the 3rd Annual Report.

Body

Task 3. Development of the statistical model to analyze the distribution of barcode abundance in tumor outgrowths.

3a. We postulated that clonal expansion of self-renewing cells can be quantified by genetic means. If intra-tumor cell population is hierarchically stratified by self-renewal and lineage potentials, then lentiviral tags integrated into self-renewing TICs will eventually predominate in tumor outgrowth population. Moreover, orthotopic transplantion in the syngeneic mouse model offers the unique chance to study tumor initiating cells during serial transplantation - the litmus test for "true" stem cells.

Since we transplant ~1000 transduced cells per immunosorted cell population per host, we necessarily undercount true unique clonal lineages when assessing outgrowth via 60mer tag abundance. It is important to recognize that the simple immunoprofiling schema merely seeks to "enrich" the TIC population, and cannot be interpreted as defining a “pure” TIC fraction. We reasoned, however, competitive repopulation with 1000 cells (and higher) per immunofraction per recipient host was sufficient to encompass cell-to-cell variation for purposes of statistical interpretation in the CD29/CD24 immunoprofile. In fact, experimental results from competitive repopulation assay using 1000 CD29+/CD24+ cells (Double Positive; DP) vs. 5000 non-DP cells, adhered well to expectation that DP-sorted fractions were indeed enriched in TICs. (Figure 1a)

We've since revised our approach based on recent published findings on tumor evolution and the genetic and epigenetic bases for intratumoral heterogeneity. In particular, while the concept of intratumoral heterogeneity has referred primarily to the immunologic diversity within a tumor, it is now important to include genomic architecture variegation as part of the discussion. (1-3). Although these studies do not directly address the finer distinctions between clonal evolution vs. cancer stem cell hypotheses, they nevertheless provide insights into intratumoral clonal

composition. In particular, we feel the concepts of "stemness" (self-renewal and lineage capacity) and repopulation kinetic needs to be decoupled and carefully investigated. Recent deep sequencing results in breast and other cancers has revealed the existence of a complex clonal architecture of both primary tumors and metastases (see below).

The advent of next-generation sequencing revealed the inadequacy of our original study design. Massively parallel sequencing (MPS) is ideally suited for querying highly complex sets of individual genetic sequences recovered from transduced tumor cells. Compared to the four individual 60mer sequences used in our studies, a modern barcoded lentiviral library created for lineage tracing analysis will need to contain >100,000 unique sequences to identify all the functional clonal lineages within a tumor population (4).

To address this concern, we devised a workaround to identify clonal lineages. Because lentiviruses integrate randomly, the flanking genomic sequence to proviral insertion sites functions *de facto* as genetic marking for the clonal lineages. We adopted the inverse PCR method to generate MPS-ready amplicons to map unique proviral insertion sites. By quantitating the number of integrants bearing 60mer tags for each immunosorted population in the competitive repopulation assay, we can then accurately assess number of lineages re-populating a tumor outgrowth.

Inverse PCR utilizes common restriction sites, typically 6bp-cutters within the lentivector and adjacent genomic sequence, followed by ligase-mediated self-circularization. PCR primer pairs originating in the lentiviral LTR, oppositely oriented with their 3' ends immediate adjacent are then used to amplify the flanking genomic insert. Due to potential GC-bias, PCR cycles are kept low (~12 cycles) but a second nested PCR reaction is carried out (~12 cycles) to further enrich bona fide proviral insertion sites. We analyzed primary tumors (outgrowths from immunosorted cells), as well as the secondary and tertiary transplants. The DNA amplicon pattern from each tumor is unique and less diverse ("thins out") over transplant passages (Figure 2).

From the DNA gels it is apparent that various tumor subclones competitively repopulate the tumor and that specific lineages are both retained and lost over transplant passages. We excised 5 bands for TOPO-TA vector cloning and subsequent Sanger sequencing. A representative “hit” is shown in Figure 3.

Sequencing of these individual amplicons was carried out on the Illumina MiSeq platform. The MiSeq can generate ~4 Gb of sequencing reads, which is sufficient to multiplex 24 tumors. We are currently analyzing the sequencing output.

Task 4. Assessment of drug- and radiation sensitivity of lentiviral barcoded tumor transplants (months 8-12).

4a. Optimization of ionizing radiation dosages needed (months 8-12). To test the competitive ecosystem of clonal populations, we subjected secondary tumor outgrowths to 6 Gy radiation then harvested genomic DNA for qPCR after 5 days. No candidate cyotoxic drugs were tested. Recovery window post-radiation was set at 5 days allowing tumor cells to undergo apoptosis and clearance. Although overt radio-resistance is not included as part of cancer stem cell hypothesis canon, lineage tracing analysis by genetic tag recovery is ideally suited to investigate this property.

We irradiated secondary and tertiary tumor outgrowths because it was necessary to have proper reference of extant 60mer sequence tags and their relative abundance. We expected 60mer sequence tags marking non-DP cells to decrease over transplant passages, but were surprised to find tags for both DP cells and non-DP cells retained even at the tertiary transplant passage (Figure 4a). No definitive conclusion could be drawn from comparison of inverse PCR-derived amplicons (Figure 4b) with qPCR quantitation data, since the sequence identity of each inverse PCR amplicon cannot be assigned to the exact 60mer sequence tag.

Therefore, we have also included the irradiated tumor samples as part our sequencing experiment and are currently analyzing the sequencing output.

Training
The scientific training at the Baylor College of Medicine affords me the opportunity to attend numerous seminars on cancer-related topics on a weekly basis. In addition to the talks held at Baylor College of Medicine, several major institutions such as MD Anderson Cancer Center, UT Health Sciences Center and The Methodist Research Institute are all within walking distance. Of note, the Lester & Sue Smith Breast Center hosts the Breast Disease Research seminar series which I attend weekly. Over this training period, I've also attended the 5th Annual Duncan Breast Center Retreat (September 2011) and also the AACR 2012 meeting in Chicago, IL (April 2012)

Key Research Accomplishments
- Selected as AACR-Aflac Scholar-In-Training Award for AACR 2012

Reportable Outcomes
- None to include; manuscript in preparation

Conclusion
In conclusion, experiments proposed in the original award for months 25-36 have been executed in timely fashion. The fellowship award was to provide funding to develop a novel strategy for investigation of the cancer stem cell, and we have responded by developing novel experimental techniques to investigate cancer stem cell hypothesis beyond the scope of the original grant proposal.

References
1. Gerlinger et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med. 2012 Mar 8;366(10):883-92. PMID: 22397650

2. Navin et al., *Tumour evolution inferred by single-cell sequencing*. Nature. 2011 Apr 7;472(7341):90-4. PMID: 21399628

3. Nik-Zainal et al., *The life history of 21 breast cancers.* Cell. 2012 May 25;149(5):994-1007. PMID: 22608083

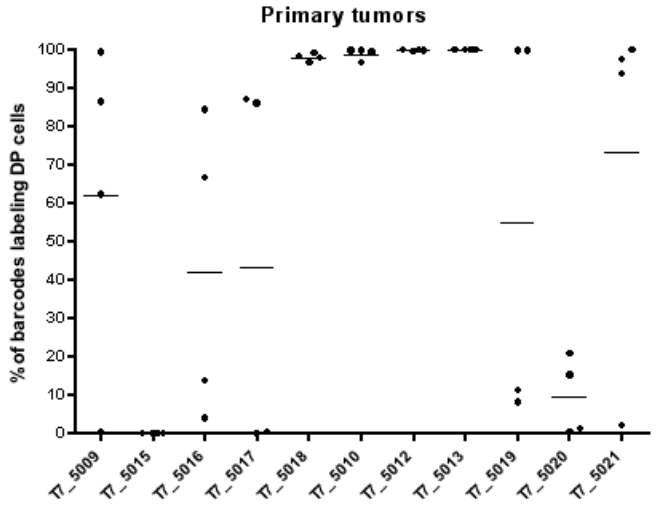
4. Lu et al., *Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding.* Nat Biotechnol. 2011 Oct 2;29(10):928-33. PMID: 21964413

Appendices
-None to include

Supplemental Data

Figure 1. (a) Relative abundance of 60mer sequence tags in competitive repopulation assays using 1000 CD29+/CD24+ cells (double-positive; DP) and 5000 non-DP cells. The y-axis is expressed as the percentage of the "DP-specific" 60mer tag within all sequence tags recovered from primary tumor transplant outgrowth. (b) Recovered 60mer sequence tags collected from secondary tumor outgrowths (#6976-6990) from tumor grafts of 5009 (primary). Y-axis plots the percentage of DP-specific 60mer sequence tags.

a.



b.

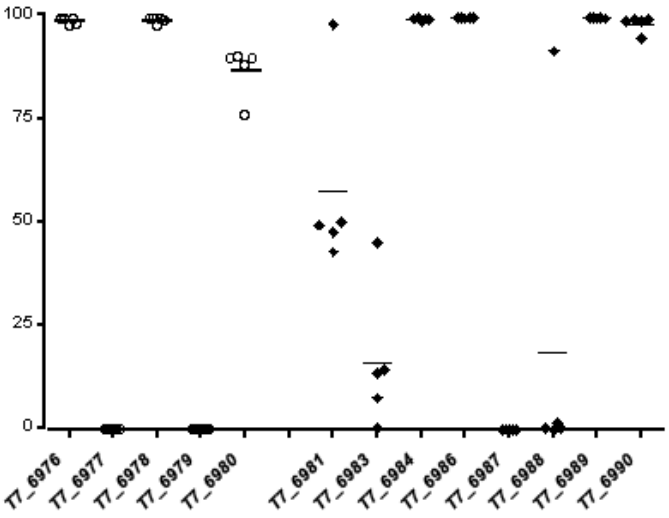


Figure 2. Inverse PCR for gDNA samples collected from resected tumor pieces of primary tumor transplant (5009) and its secondary transplant outgrowth (6983)

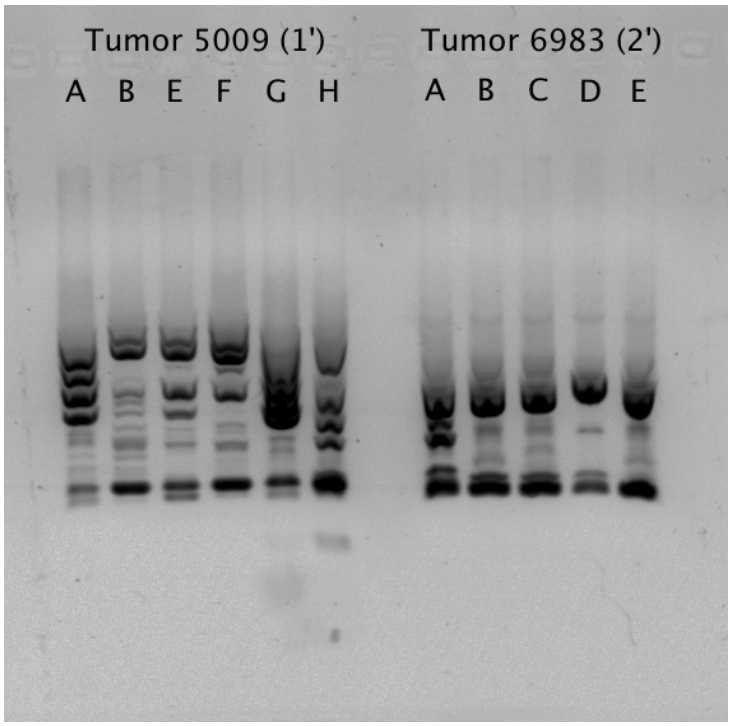


Figure 3. Sanger sequencing result of an excised PCR band cloned into TOPO-TA and BLAT'ed on UCSC Genome Browser. The top result is 100% and spans entire DNA amplicon sequence

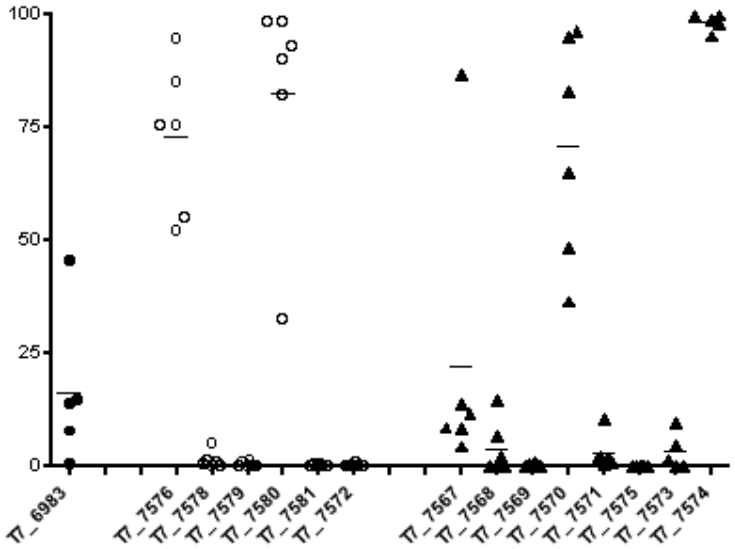
Mouse BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	699	1	700	702	100.0%	10	+	80269485	80270185	701
browser details	YourSeq	265	171	615	702	94.0%	10	-	128177668	128178255	588
browser details	YourSeq	264	172	481	702	94.7%	15	+	76570385	76864599	294215
browser details	YourSeq	248	173	528	702	87.1%	19	-	5978542	5978866	325
browser details	YourSeq	229	172	482	702	89.1%	2	-	170081959	170082256	298
browser details	YourSeq	228	172	479	702	91.2%	5	-	105864902	105865207	306
browser details	YourSeq	227	171	528	702	93.9%	13	+	3548237	3663866	115630
browser details	YourSeq	224	201	536	702	88.6%	2	-	154527857	154528184	328

Figure 4. (**a**) qPCR quantitation of 60mer sequence tag abundance in tertiary transplant outgrowths. Y-axis is the percentage of DP-specific tag (DP cells from the original immunosort for primary tumor transplant). Tertiary tumors sham group (open circle) and 6 Gy irradiated (filled triangle) descended from transplant lineage of 5009 (primary) -> 6983 (secondary). (**b**) Inverse PCR amplicons for sham treated (7572, 7575-7581; tertiary) and 6 Gy irradiated tumor samples (7567-7571, 7573, 7574; tertiary)

a.



b.